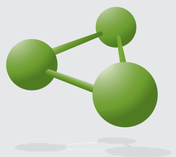
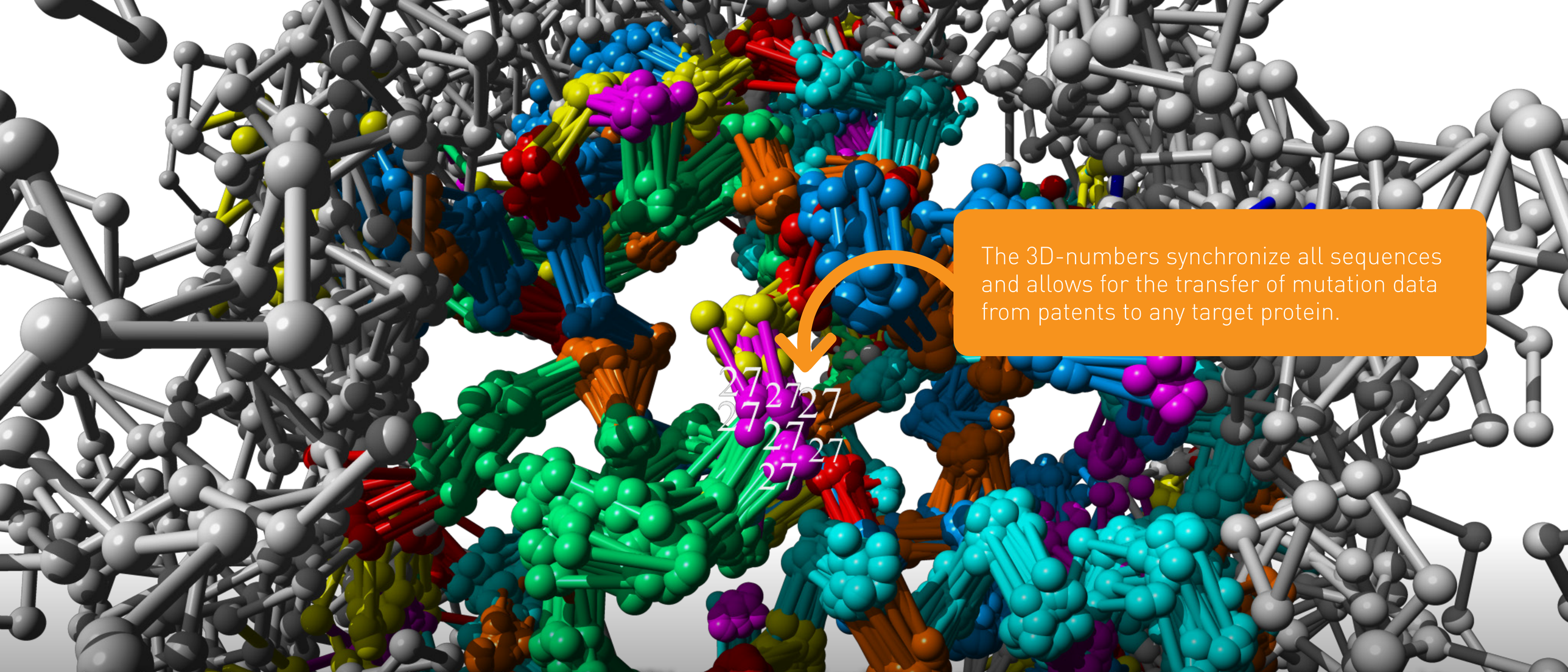
3DM'S PATENT ANALYSIS SUITE BY BIO-PRODICT

1. 3DM introduction
2. 3DM's patent analysis suite

bio-prodict
protein predictions

1. 3DM INTRODUCTION



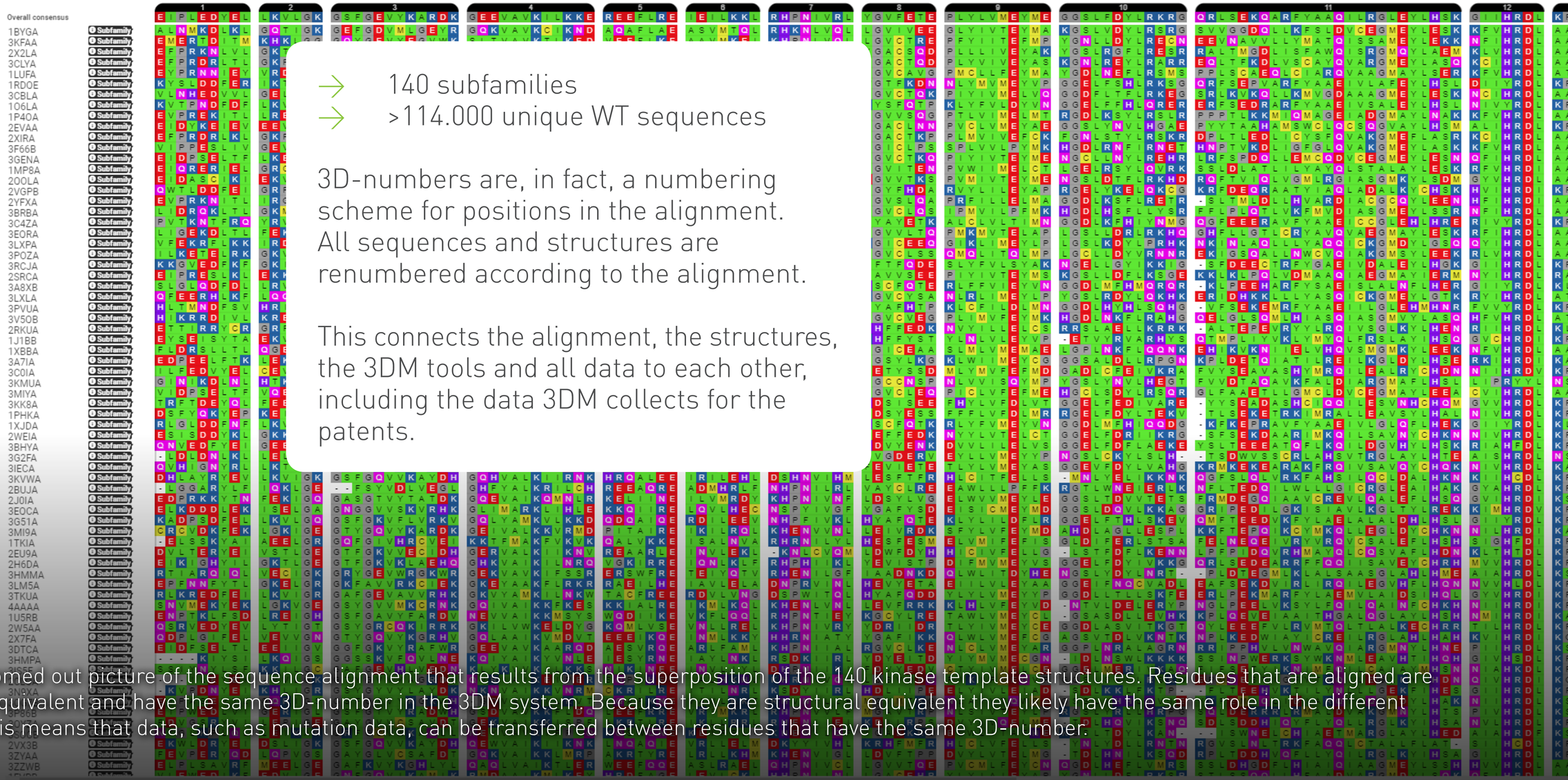


The 3D-numbers synchronize all sequences and allows for the transfer of mutation data from patents to any target protein.

In the first step of the generation of a 3DM superfamily platform 3DM superimposes all available structures for the protein family and groups the structures into subfamilies. From these groups one template structure is selected that is used to align sequences for which no structures are available thereby forming subfamily alignments. Here 140 of such template structures are shown for the kinase protein family. The resulting 140 subfamilies are combined into one large superfamily alignment. In total the alignment for the kinases contains over 114.000 WT sequences. A 3D-numbering scheme, in which all structural equivalent residues get the same number (e.g. are in the same column in the alignment), is applied to all sequences, all structures, and to the alignments. Here, the labels of the purple residues located at 3D-position 27 are shown. The 3D-numbers synchronize all sequences and structures.

Dashboard
3DM System
Start
Alignment
Alignment statistics
Correlated mutations
Visualize
Phylogeny
Panel design
Patents
Proteins
Search
System

Alignment (structures excluded)



- 140 subfamilies
- >114.000 unique WT sequences

3D-numbers are, in fact, a numbering scheme for positions in the alignment. All sequences and structures are renumbered according to the alignment.

This connects the alignment, the structures, the 3DM tools and all data to each other, including the data 3DM collects for the patents.

This is a zoomed out picture of the sequence alignment that results from the superposition of the 140 kinase template structures. Residues that are aligned are structural equivalent and have the same 3D-number in the 3DM system. Because they are structural equivalent they likely have the same role in the different proteins. This means that data, such as mutation data, can be transferred between residues that have the same 3D-number.

MANY DIFFERENT DATA TYPES ARE COLLECTED FOR ALL SEQUENCES IN THESE LARGE ALIGNMENTS

→ **Mutation data**
From literature: 264.605...
From patents: 1.683 mutations have been extrated from claimes.

→ **Structure data**
PPI, ligand contacts, bridges, solvent accessibility, flexibility/RMSD, ...

→ **Alignment data**
Conservation, correlated mutations, family specific residues, ...

→ **SNP data**
SNP databases, genome sequencing projects

→ **Data- and tool integration**
All data are stored connected to the 3D numbering scheme and thus to the alignment, to the structures, all 3DM tools, and each other.

3DM collects many different data types from all kinds of sources for all the 114.000 proteins that are available in this kinase protein family. These are all stored connected to the 3D-numbers. This connects all data to each other making it possible to make very complex queries. For example, it now takes a couple of seconds to find out if mutations reported in the literature that cause effects on specificity are more often found at positions that show a correlated mutation behavior in the alignment. Manually such comparisons can easily be a year work.

2. 3DM'S PATENT ANALYSIS SUITE

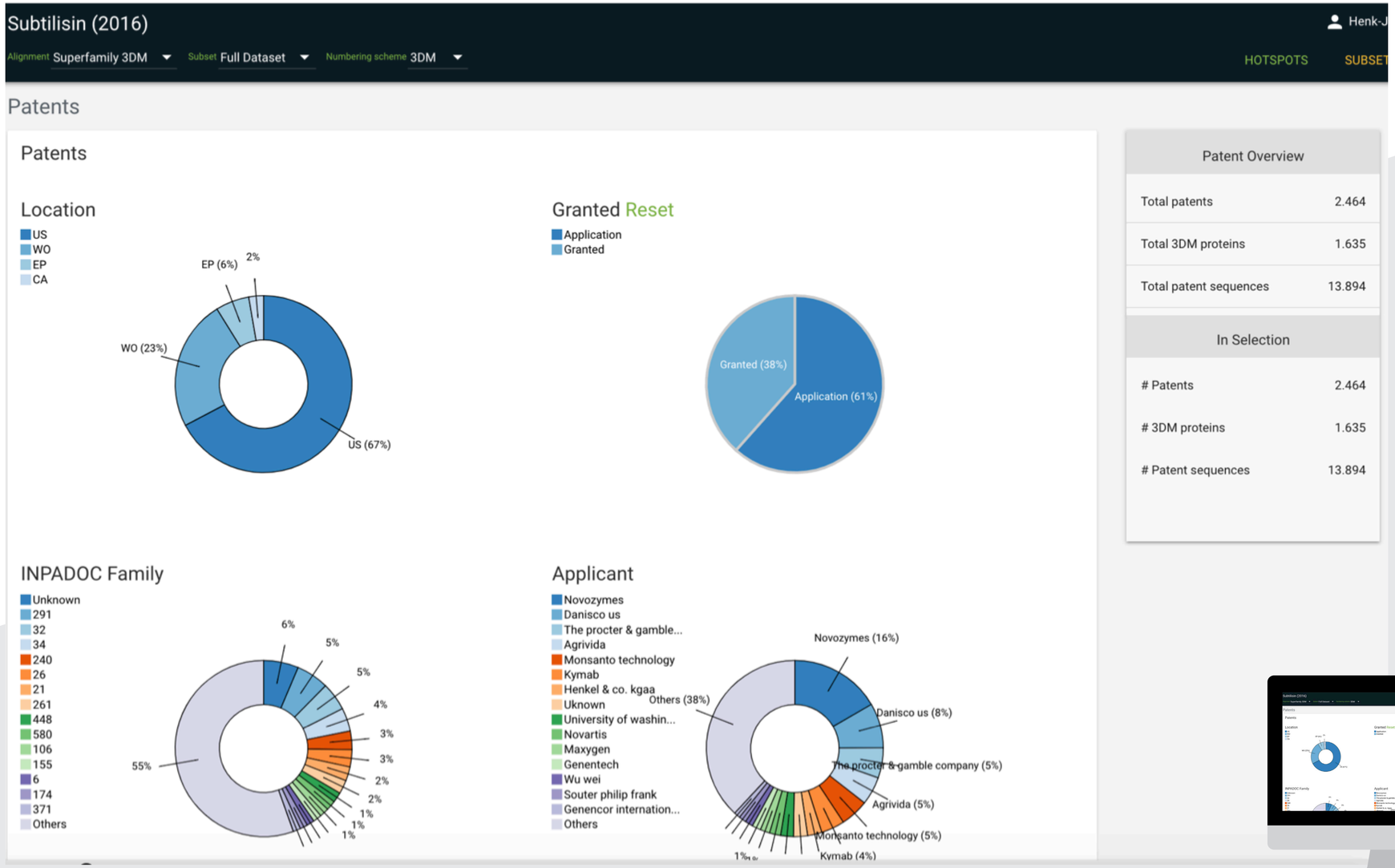
PATENT DATA EXTRACTION

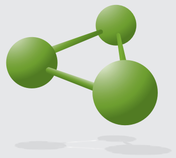
→ **Collection of patents for complete protein families**

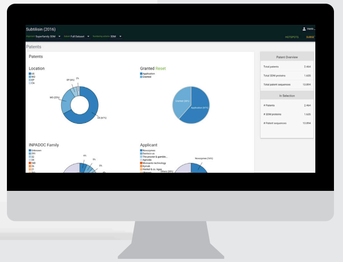
- Claimed and mentioned sequences
- Claimed sequence spaces
- Claimed mutations › Patent family information
- Bibliographic data:
    - I Priority, filing and publication dates
    - II Applicants
    - III Titles & abstracts
    - IV Many others...
- Inventors
- Many other data types

3DM collects the patents for complete protein superfamilies by combining complex patent sequence searches with a smart patent family search. For important heavily patented protein families, such as the kinase protein family or the antibody protein family, 3DM collects many (7.944 and 44.859 respectively) patents. From these patents many different data types are extracted, such as claimed sequences and claimed mutations. For the kinase and antibody protein families 3DM extracted 131.578 and 703.356 sequences, and 1.683 and 5.180 mutations, respectively.

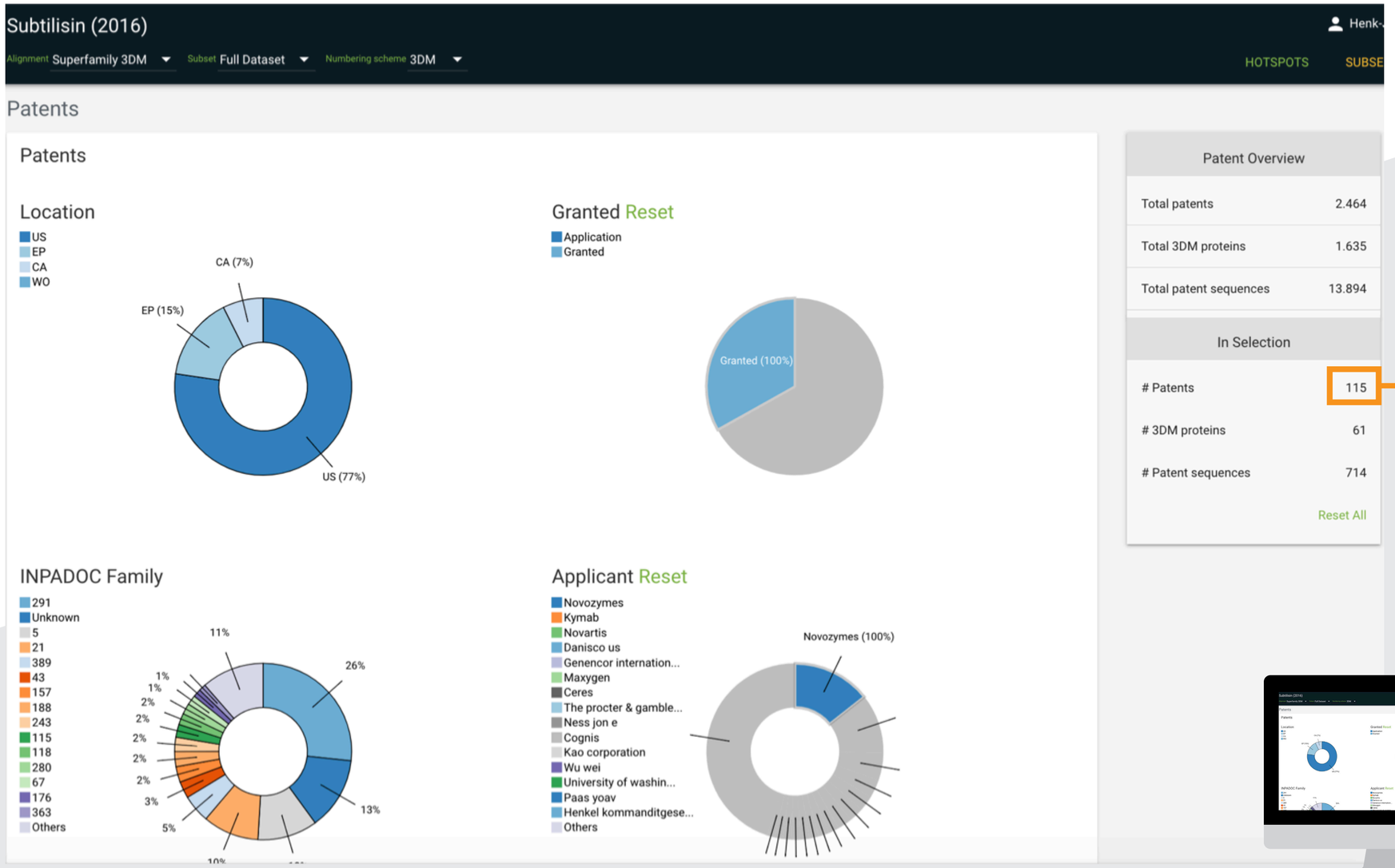
DIFFERENT ONLINE SELECTION TOOLS FOR ANALYSIS OF THE PATENTS



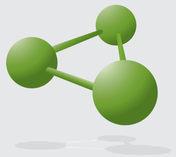


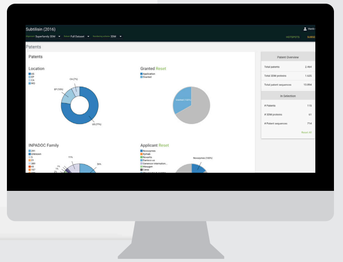


DIFFERENT ONLINE SELECTION TOOLS FOR ANALYSIS OF THE PATENTS

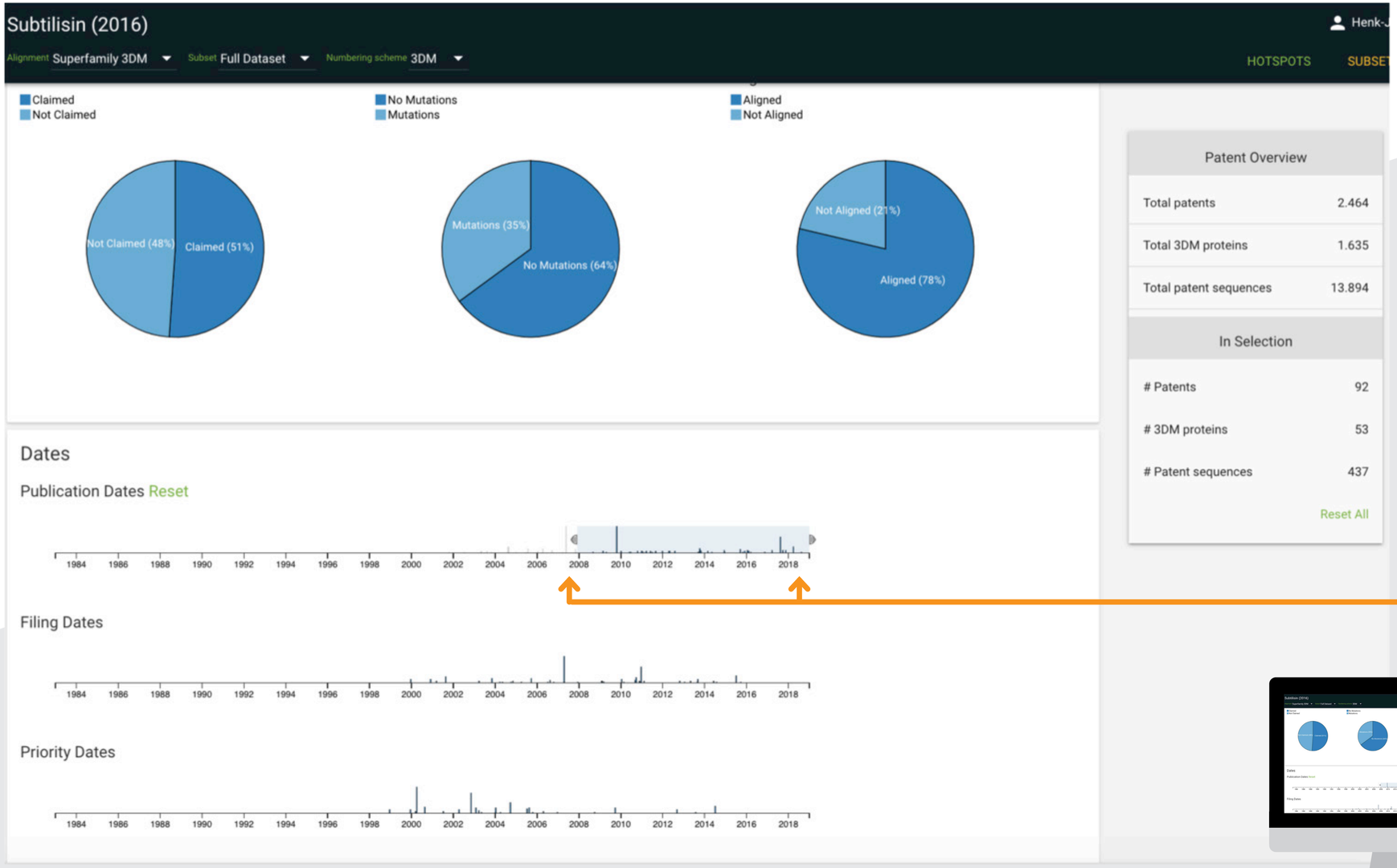


Here all granted patents from Novozymes are selected. Novozymes has 115 granted patents in this protein family.

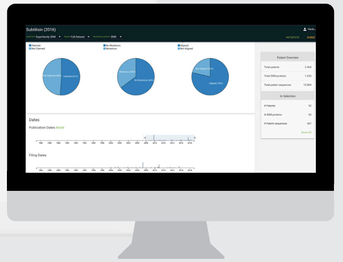




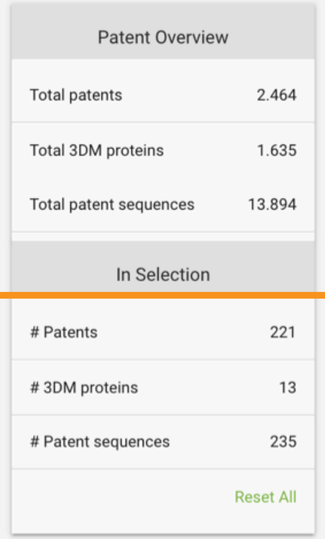
DIFFERENT ONLINE SELECTION TOOLS FOR ANALYSIS OF THE PATENTS



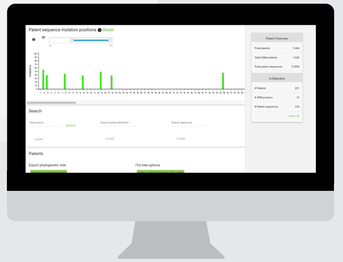
Of which 92 have been published since 2008



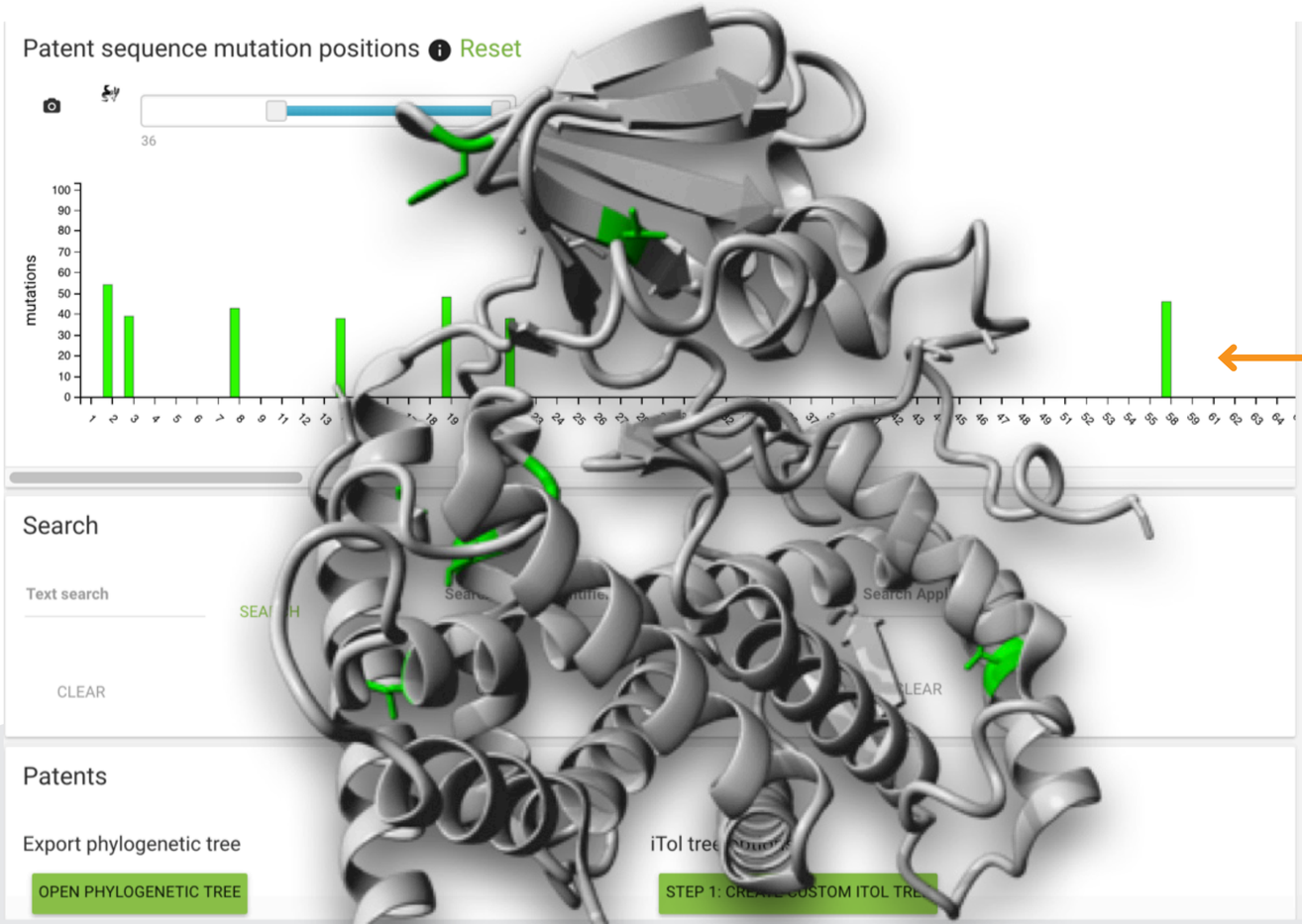
DIFFERENT VISUALISATION OPTIONS



Interactive histogram visualising mutations claimed by novozymes.



MANY OPTIONS FOR VISUALIZING PATENT DATA



Patent Overview	
Total patents	2.464
Total 3DM proteins	1.635
Total patent sequences	13.894

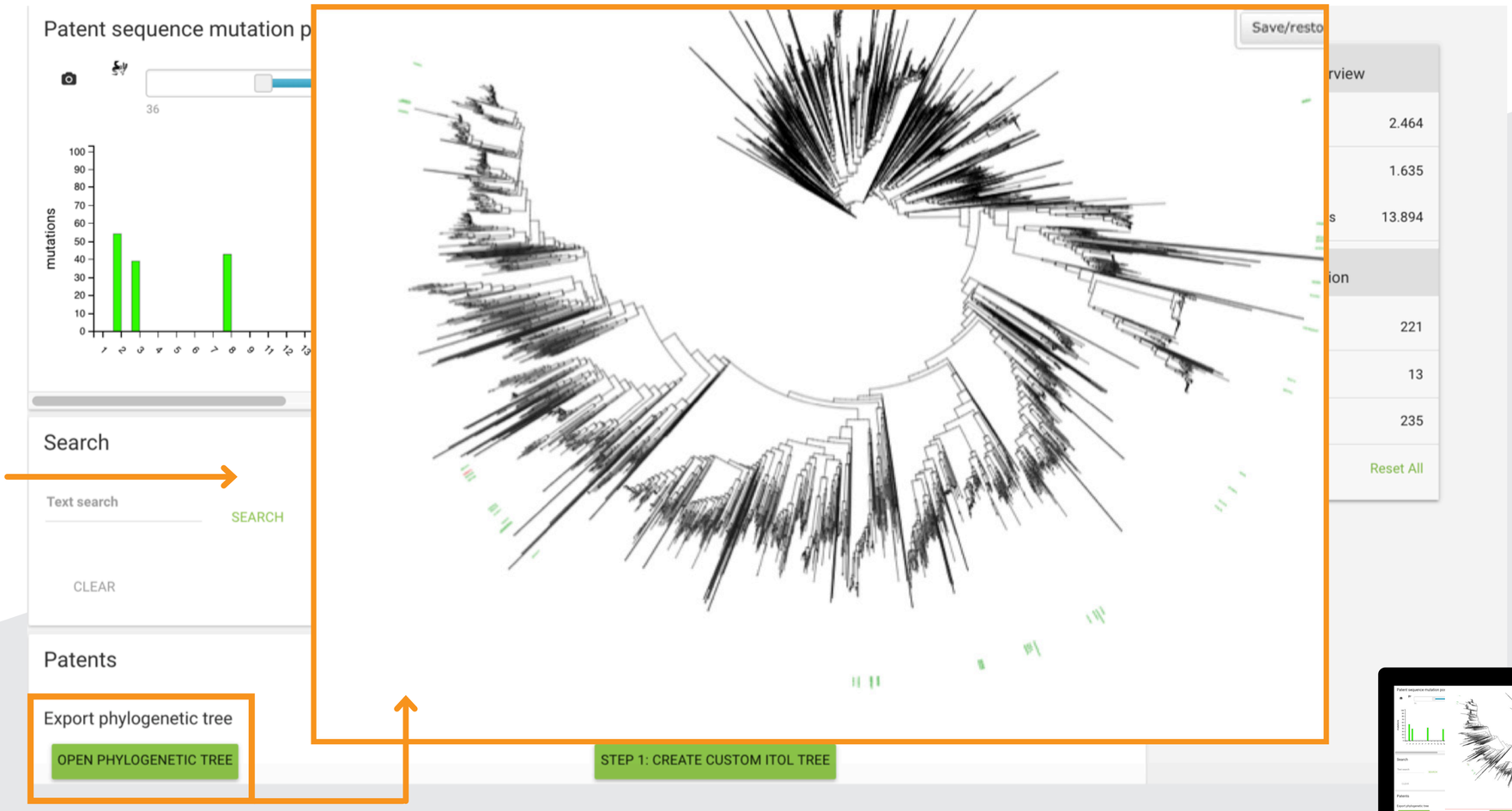
In Selection	
# Patents	221
# 3DM proteins	13
# Patent sequences	235

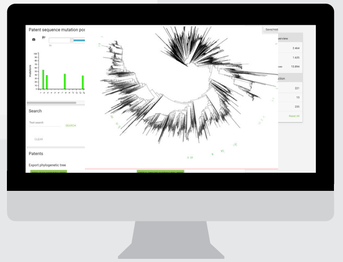
Reset All

The selected mutation data can instantly be visualized in any available protein structure or homology model.

MANY OPTIONS FOR VISUALIZING PATENT DATA

The patented sequences can instantly be visualized in the phylogenetic tree of the superfamily for FTO analysis.





SUMMARY OF THE SELECTED PATENTS IS PROVIDED

Link to interactive analysis tool for each patent

Patent List

DOWNLOAD PATENT SEQUENCES

A list of the selected patents is provided.

First | Previous | 1 | 2 | Next | Last

92 patents in selection

Family:115 - US7951573 B2

External link

Subtilases

Applicant: Novozymes

Publication type: Granted

Sequences: 8

Published: 31/05/2011
Filed: 18/11/2010
Priority: 16/08/2005

Abstract

The present invention relates to novel subtilases from wild-type strains of Bacillus, especially the Bacillus strains ZI344, EP655, P203, EP63, ZI120, ZI130, ZI1342 and ZI140, and to methods of construction and production of these proteases. Further, the present invention relates to use of the claimed subtilases in detergents, such as a laundry detergent or an automatic dishwashing detergent.

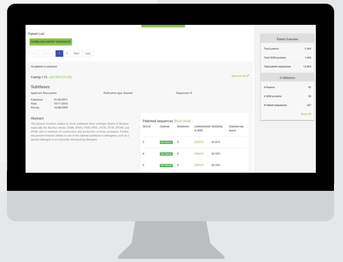
Patented sequences Show more...

SEQ ID	Claimed	Mutations	Linked protein in 3DM	Similarity	Claimed seq. space
2	Not Claimed	0	S5VEF0	81.87%	
4	Not Claimed	0	S5VEF0	82.93%	
6	Not Claimed	0	S5VEF0	82.93%	

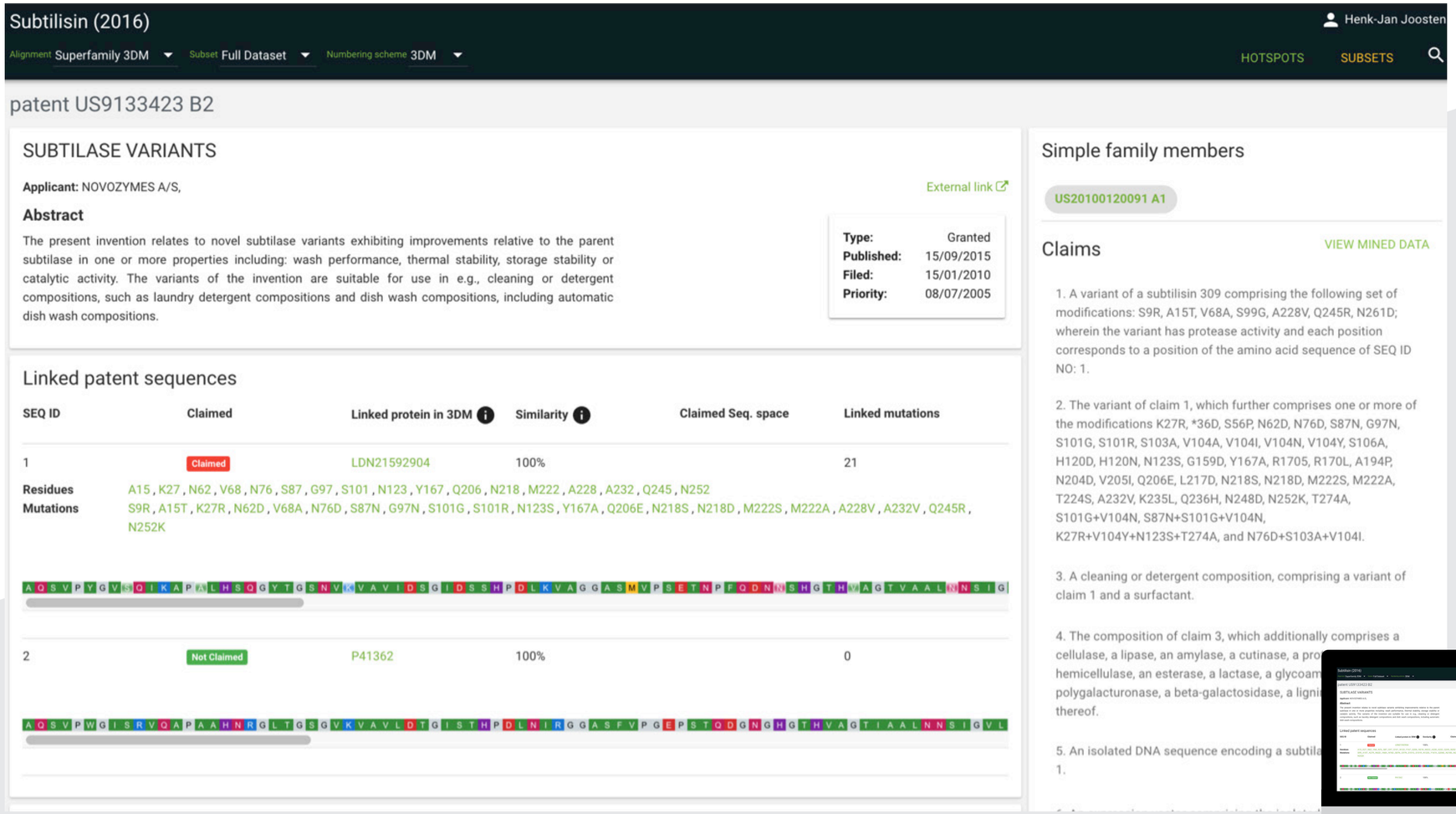
Patent Overview	
Total patents	2.464
Total 3DM proteins	1.635
Total patent sequences	13.894

In Selection	
# Patents	92
# 3DM proteins	53
# Patent sequences	437

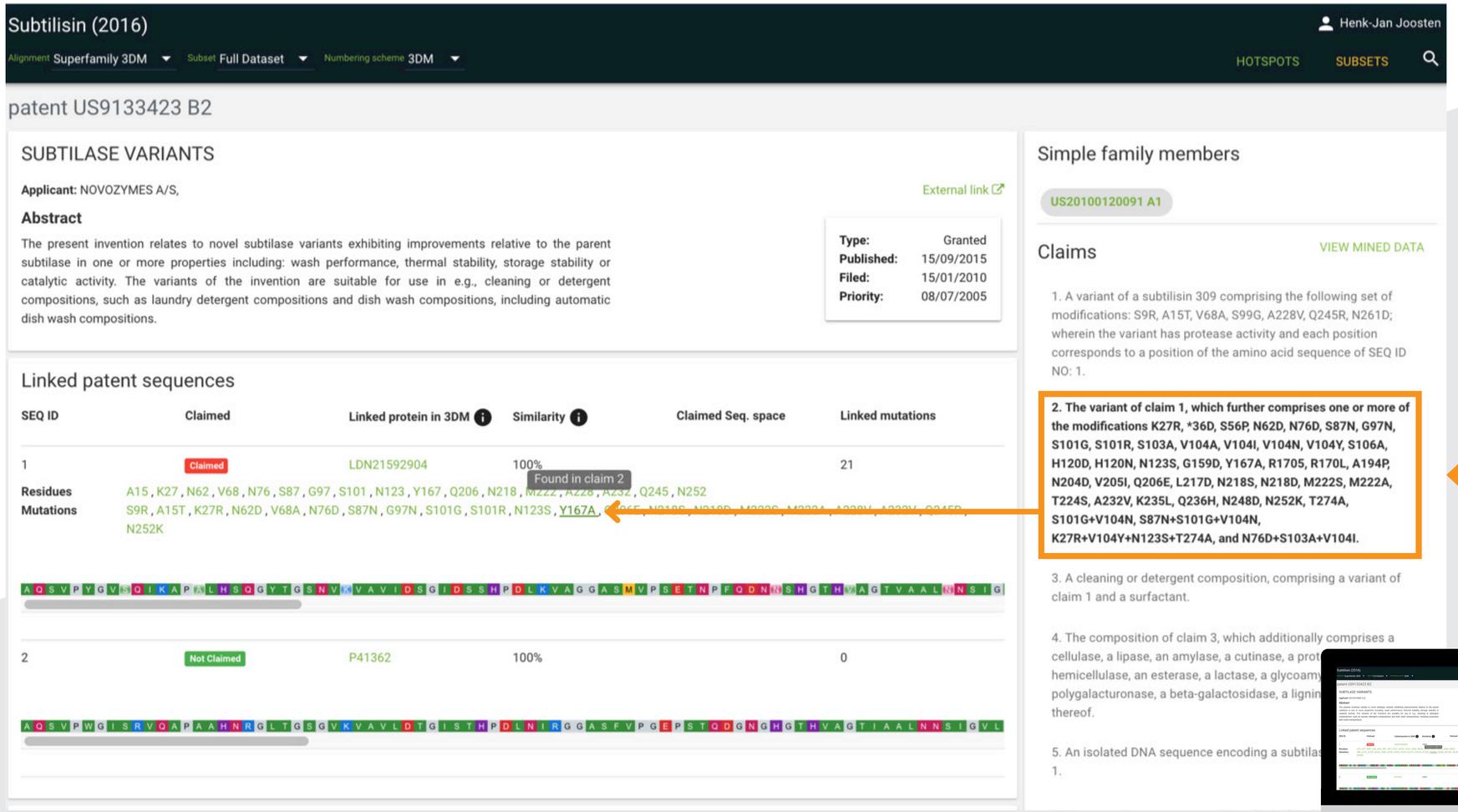
Reset All



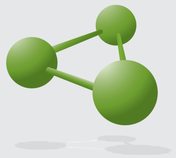
INTERACTIVE TOOL FOR ANALYSIS A PATENT SHOWING CLAIMED SEQUENCES AND MUTATIONS

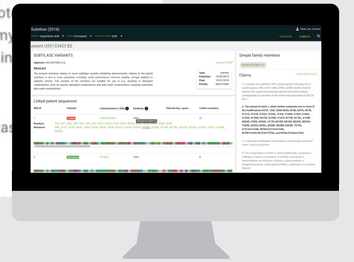


MOVING MOUSE OVER DATA WILL HIGHLIGHT THE CLAIM IN WHICH 3DM FOUND THE SELECTED DATA POINT

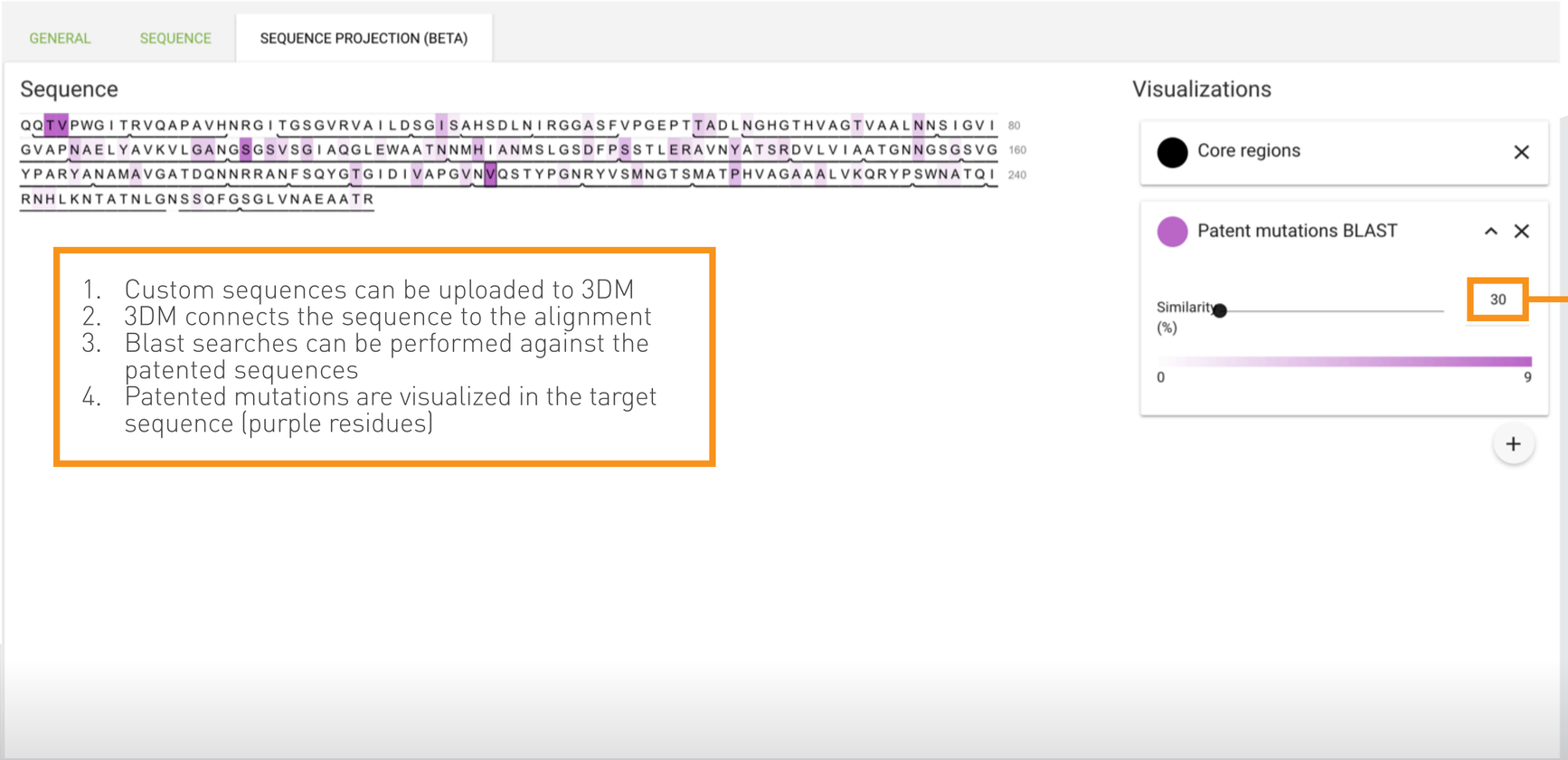


Interactively analyse claimed sequences and mutations in a patent.





TOOL FOR PROJECTING PATENT DATA ON A CUSTOM TARGET SEQUENCE

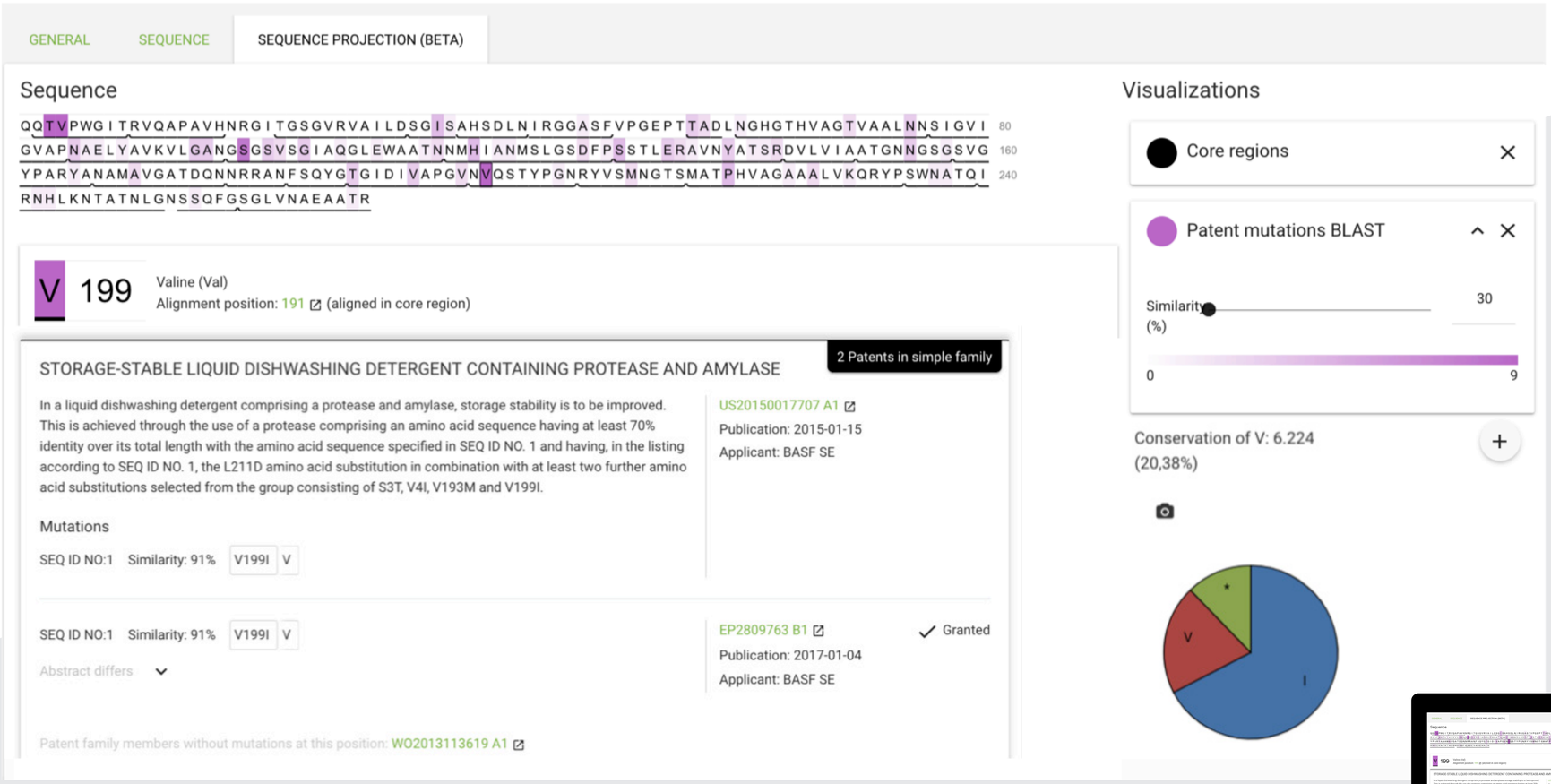


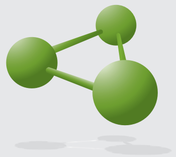
1. Custom sequences can be uploaded to 3DM
2. 3DM connects the sequence to the alignment
3. Blast searches can be performed against the patented sequences
4. Patented mutations are visualized in the target sequence (purple residues)

Here the cut-off was set on 30% sequence identity. So, all mutations claimed in sequences that are >30% identical are plotted on the target sequence.

This is a screenshot of 3DM's sequence projection tool. This tool allows for the visualization of data from homologues sequences directly on your target protein. This feature is unique to 3DM and makes 3DM such a powerful tool. As different homologues proteins have different numbering schemes this is very difficult and time consuming to do manually. For instance, if you are interested in a certain amino acid in your target, it is very difficult to find all patents that claim a mutation at that specific position. The 3D-numbering scheme synchronizes the patented sequences to any protein target making such complex searches very easy.

CLICKING ON A PURPLE RESIDUE PROVIDES A LIST OF PATENTS THAT HAVE CLAIMED MUTATIONS AT THE SELECTED POSITION



For more information visit our website:
www.bio-prodict.nl

or request information via:
info@bio-prodict.nl